New Techniques for Upper-Bounding the ML Decoding Performance of Binary Linear Codes

Xiao Ma, *Member, IEEE*, Jia Liu and Baoming Bai, *Member, IEEE*,

Abstract

In this paper, new techniques are presented to either simplify or improve most existing upper bounds on the maximum-likelihood (ML) decoding performance of the binary linear codes over additive white Gaussian noise (AWGN) channels. Firstly, the recently proposed union bound using truncated weight spectrum by Ma *et al* is re-derived in a detailed way based on Gallager's first bounding technique (GFBT), where the "good region" is specified by a sub-optimal list decoding algorithm. The error probability caused by the bad region can be upper-bounded by the tail-probability of a binomial distribution, while the error probability caused by the good region can be upper-bounded by most existing techniques. Secondly, we propose two techniques to tighten the union bound on the error probability caused by the good region. The first technique is based on pair-wise error probabilities. The second technique is based on triplet-wise error probabilities, which can be upper-bounded by the fact that any three bipolar vectors form a non-obtuse triangle. The proposed bounds improve the conventional union bounds but have a similar complexity since they involve only the Q-function. The proposed bounds can also be adapted to bit-error probabilities.

Index Terms

This work was presented in part at ISIT'2011.

X. Ma and J. Liu are with the Department of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou 510275, China. (E-mail: maxiao@mail.sysu.edu.cn)

B. Bai is with State Key Lab. of ISN, Xidian University, Xi'an 710071, China. (E-mail: bmbai@mail.xidian.edu.cn)

Additive white Gaussian noise (AWGN) channel, binary linear block code, Gallager's first bounding technique (GFBT), list decoding, maximum-likelihood (ML) decoding, union bound.

I. INTRODUCTION

In most scenarios, there do not exist easy ways to compute the exact decoding error probabilities for specific codes and ensembles. Therefore, deriving tight analytical bounds is an important research subject in the field of coding theory and practice. Since the early 1990s, spurred by the successes of the near-capacity-achieving codes, renewed attentions have been paid to the performance analysis of the maximum-likelihood (ML) decoding algorithm. Though the ML decoding algorithm is prohibitively complex for most practical codes, tight bounds can be used to predict their performance without resorting to computer simulations. As shown in [1][2], most bounding techniques have connections to either the 1965 Gallager bound [3–6] or the 1961 Gallager-Fano bound [7–18]. This paper is relevant to the 1961 Gallager-Fano bound, which is also called Gallager's first bounding technique (GFBT) in the literature. Our efforts focus on tightening the simplest conventional union bound, which is simple but loose and even diverges in the low-SNR region. Similar to many previously reported upper bounds surveyed in [2], our basic approach is based on GFBT

$$\begin{aligned} \Pr\{E\} &= \Pr\{E, \underline{y} \in \mathcal{R}\} + \Pr\{E, \underline{y} \notin \mathcal{R}\} && (1) \\ &\leq \Pr\{E, \underline{y} \in \mathcal{R}\} + \Pr\{\underline{y} \notin \mathcal{R}\}, && (2) \end{aligned}$$

where E denotes the error event, $\underline{y}$ denotes the received signal vector, and $\mathcal{R}$ denotes an arbitrary region around the transmitted signal vector which is usually interpreted as the "good region". As pointed out in [2], the choice of the region $\mathcal{R}$ is very significant, and different choices of this region have resulted in various different improved upper bounds. Intuitively, the more similar the region $\mathcal{R}$ is to the Voronoi region of the transmitted codeword, the tighter the upper bound is. However, most existing improved upper bounds have higher computational complexity than

the conventional union bound.

Different from most of the existing works, we define the good region using a list decoding algorithm. The basic idea is as follows. Upper bounds on the word-error probability for the list decoding algorithm (which is suboptimal) can also be applied to an ML decoding algorithm, while the list decoding algorithm can limit competitive candidate codewords.

Structure: The rest of this paper is organized as follows. In Sec. II, we present an upper bound of the angle formed by any three bipolar vectors, which will be used to upper-bound the triplet-wise error probabilities. In Sec. III, we re-derive, in a detailed way within the framework of the GFBT, the recently proposed union bound using truncated weight spectrum by Ma *et al* [19]. On one hand, the truncation technique is helpful when the whole weight spectrum is unknown or not computable. On the other hand, the truncation technique can be combined with any other upper-bounding techniques, potentially resulting in tighter upper bounds. In Sec. IV, we propose two techniques to improve the union bound. The first technique is based on the pair-wise error probabilities, which can be tightened by employing the independence of the error event and certain components of the received random vectors. The second technique is based on the triplet-wise error probabilities, which is shown to be a non-decreasing function of the angle formed by the transmitted codeword and the other two codewords. In Sec. V, the proposed bounds are adapted to ensembles of codes and bit-error probabilities. Numerical examples are provided in Sec. VI and we conclude this paper in Sec. VII.

II. PRELIMINARIES

A. Geometrical Properties of Binary Codes

Let $\mathbb{F}_2 = \{0, 1\}$ and $\mathcal{A}_2 = \{-1, +1\}$ be the binary field and the bipolar signal set, respectively. We use $W_H(\underline{v})$ to denote the Hamming weight of a binary vector $\underline{v} \stackrel{\Delta}{=} (v_0, v_1, \cdots, v_{n-1}) \in \mathbb{F}_2^n$. We use $\|\underline{y}\|$ to denote the magnitude of a real vector $\underline{y} \stackrel{\Delta}{=} (y_0, y_1, \cdots, y_{n-1}) \in \mathbb{R}^n$, that is, $\|\underline{y}\| = \sqrt{\sum_{0 \le t < n} y_t^2}$. Let $\mathcal{C}[n, k]$ be a binary linear block code of dimension k and length n with

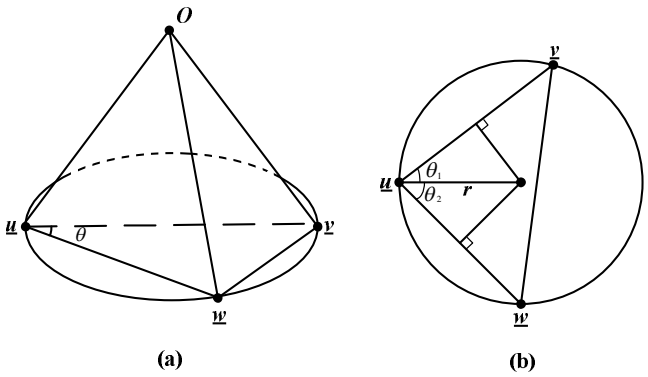


Fig. 1. Geometrical representation of three bipolar vectors.

a generator matrix G of size $k \times n$, that is,

$$\mathcal{C} \triangleq \left\{ \underline{c} \in \mathbb{F}_2^n \mid \underline{c} = \underline{u}G, \underline{u} \in \mathbb{F}_2^k \right\}. \tag{3}$$

Let $A_{i,j}$ denote the number of codewords $\underline{c} = \underline{u}G$ with $W_H(\underline{u}) = i$ and $W_H(\underline{c}) = j$. Then $\{A_j \triangleq \sum_i A_{i,j},\ 0 \le j \le n\}$ is referred to as the weight spectrum of the given code $\mathcal{C}$.

Consider the binary phase shift keying (BPSK) mapping $\phi : \mathbb{F}_2^n \mapsto \mathcal{A}_2^n$ taking $\underline{s} = \phi(\underline{v})$ by $s_t = 1 - 2v_t$ for $0 \le t \le n-1$. The image of $\mathcal{C}$ under this mapping is denoted by $\mathcal{S} \triangleq \phi(\mathcal{C})$. Hereafter, we may not distinguish $\underline{c} \in \mathcal{C}$ from its image $\underline{s} \in \mathcal{S}$ when representing a codeword. Let $d_H(\underline{v}^{(1)}, \underline{v}^{(2)}) \triangleq W_H(\underline{v}^{(1)} - \underline{v}^{(2)})$ be the Hamming distance between two binary vectors $\underline{v}^{(1)}$ and $\underline{v}^{(2)}$. Then their Euclidean distance $\|\phi(\underline{v}^{(1)}) - \phi(\underline{v}^{(2)})\|$ is equal to $2\sqrt{d_H(\underline{v}^{(1)}, \underline{v}^{(2)})}$. Obviously, the vectors in $\mathcal{A}_2^n$ (hence the bipolar codewords) are distributed on an n-dimensional sphere of radius $\sqrt{n}$ centered at the origin O of $\mathbb{R}^n$. We have the following lemma.

Lemma 1: Let $\underline{u}$, $\underline{v}$ and $\underline{w}$ be three bipolar vectors of length n. Let θ be the angle formed by the two vectors $\overrightarrow{uv} \triangleq \underline{v} - \underline{u}$ and $\overrightarrow{uw}$. Then we have

$$\theta \le \min\left\{ \frac{\pi}{2},\ \arccos\sqrt{\frac{d_1}{n}} + \arccos\sqrt{\frac{d_2}{n}} \right\}, \tag{4}$$

where $d_1 = d_H(\underline{u}, \underline{v})$ and $d_2 = d_H(\underline{u}, \underline{w})$.

Proof: To make the proof more readable, we have drawn the three bipolar vectors in a three-dimensional space, as shown in Fig. 1 (a). In essence, with a properly chosen orthogonal transformation, the three vectors can be viewed as three points in $\mathbb{R}^3$ (a three-dimensional subspace of $\mathbb{R}^n$). It should be noted that orthogonal transformations preserve inner products and (hence) lengths as well as angles.

It has been pointed out in [20] (without proof) that any three bipolar vectors form a non-obtuse triangle, which means $\theta \leq \pi/2$. For completeness, we re-derive this bound in a detailed way. Let θ be the angle formed by $\overrightarrow{uv}$ and $\overrightarrow{uw}$. It suffices to prove that the inner product $\overrightarrow{uv} \cdot \overrightarrow{uw}$ is non-negative. Actually, if $v_t \neq w_t$, $(v_t - u_t)(w_t - u_t) = 0$ since either $v_t = u_t$ or $w_t = u_t$ must hold; if $v_t = w_t$, $(v_t - u_t)(w_t - u_t) \geq 0$. Therefore

$$\overrightarrow{uv} \cdot \overrightarrow{uw} = \sum_t (v_t - u_t)(w_t - u_t) \geq 0. \tag{5}$$

To complete the proof of this lemma, consider the circumscribed circle of the triangle formed by the three points $\underline{u}$, $\underline{v}$ and $\underline{w}$ (Fig. 1 (b)). Let r be its radius. The angle can be written as $\theta = \theta_1 + \theta_2$, where $\cos\theta_1 = \|\overrightarrow{uv}\|/(2r)$ and $\cos\theta_2 = \|\overrightarrow{uw}\|/(2r)$. It is then not difficult to verify that

$$\theta = \arccos\frac{\sqrt{d_1}}{r} + \arccos\frac{\sqrt{d_2}}{r}. \tag{6}$$

Noticing that the right hand side (RHS) of (6) is increasing with r and that $r \leq \sqrt{n}$, we have

$$\theta \leq \arccos\sqrt{\frac{d_1}{n}} + \arccos\sqrt{\frac{d_2}{n}}. \tag{7}$$

■

B. Union Bounds

Let $\underline{c} = (c_0, c_1, \cdots, c_{n-1}) \in \mathcal{C}$ be a codeword. Suppose that $\underline{s} = \phi(\underline{c})$ is transmitted over an AWGN channel. Let $\underline{y} = \underline{s} + \underline{z}$ be the received vector, where $\underline{z}$ is a vector of independent Gaussian random variables with zero mean and variance σ^2. For AWGN channels, the ML decoding is

equivalent to finding the nearest signal vector $\underline{\hat{s}} \in \mathcal{S}$ to $\underline{y}$. A decoding error occurs whenever $\underline{\hat{s}} \neq \underline{s}$. Let E be the decoding error event (under ML decoding). Generally, it is a difficult task to calculate the decoding error probability $\Pr\{E\}$. Hence one usually turns to bounding techniques. Due to the symmetry of the channel and the linearity of the code, the conditional error probability does not depend on the transmitted codeword, see, e.g., [21]. Therefore, without loss of generality, we assume that the all-zero codeword $\underline{c}^{(0)}$ is transmitted. The simplest upper bound is the union bound

$$\begin{aligned} \Pr\{E\} &= \Pr\left\{\bigcup_d E_d\right\} \\ &\leq \sum_d \Pr\{E_d\} \\ &\leq \sum_d A_d Q\left(\frac{\sqrt{d}}{\sigma}\right), \end{aligned} \tag{8}$$

where E_d is the event that there exists at least one codeword of Hamming weight $d \geq 1$ that is nearer than $\underline{c}^{(0)}$ to $\underline{y}$, and $Q\left(\frac{\sqrt{d}}{\sigma}\right)$ is the pair-wise error probability with

$$Q(x) \triangleq \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, \mathrm{d}z. \tag{9}$$

The question is, how many terms do we need to count for the summation in the above bound? If too few terms are counted, we will obtain a lower bound of the upper bound, which may be neither an upper bound nor a lower bound; if too many are counted, we need pay more efforts to compute the distance distribution and only a loose upper bound will be obtained. To get a tight upper bound, we may determine the terms by analyzing the facets of the Voronoi region of the codeword $\underline{c}^{(0)}$ [22] [20], which is a difficult task for a general code.

It is well-known that the conventional union bound is loose and even diverges (≥ 1) in the low-SNR region. One objective of this paper is, without too much complexity increase, to reduce the number of involved terms in the conventional union bound. The other objective of this paper is to tighten the bound on $\Pr\{E_d\}$, which used to be upper-bounded by the pair-wise error

probability, where intersections of half-spaces related to codewords other than the transmitted one are counted more than once. For some of well-known existing improved bounds based on GFBT, such as the sphere bound (SB), the tangential-sphere bound (TSB) and the Divsalar bound, see the monograph [2, Ch. 3] and the references therein.

III. Upper Bounds Using Truncated Weight Spectrum

Recently, Ma *et al* [19] proposed a union bound which involves only truncated weight spectrum. In this section, we re-derive this "truncated" union bound within the framework of GFBT, where the region $\mathcal{R}$ is defined in an unusual way based on the following conceptual suboptimal list decoding algorithm.

***Algorithm** 1:* (A list decoding algorithm for the purpose of performance analysis)

S1. Make hard decisions, i.e., for $0 \leq t \leq n-1$,

$$\hat{y}_t = \begin{cases} 0, & y_t > 0 \\ \\ 1, & y_t \leq 0 \end{cases} . \tag{10}$$

Then the channel $c_t \rightarrow \hat{y}_t$ becomes a memoryless binary symmetric channel (BSC) with cross probability $p_b \stackrel{\Delta}{=} Q\left(\frac{1}{\sigma}\right)$.

S2. List all codewords within the Hamming sphere with center at $\underline{\hat{y}}$ of radius $d^* \geq 0$. The resulting list is denoted as $\mathcal{L}_{\underline{y}}$.

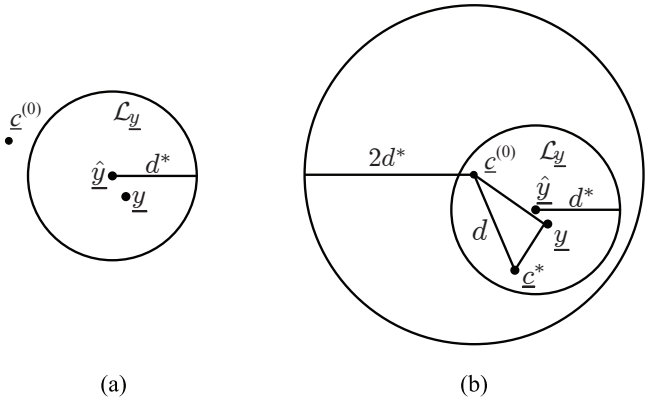
S3. If $\mathcal{L}_{\underline{y}}$ is empty, declare a decoding error; otherwise, find the codeword $\underline{c}^* \in \mathcal{L}_{\underline{y}}$ such that $\phi(\underline{c}^*) \in \mathcal{S}$ is closest to $\underline{y}$.

❑

Now we define

$$\mathcal{R} \stackrel{\Delta}{=} \left\{ \underline{y} | \underline{c}^{(0)} \in \mathcal{L}_{\underline{y}} \right\}. \tag{11}$$

In words, the region $\mathcal{R}$ consists of all those $\underline{y}$ having at most d^* non-positive components. The decoding error occurs in two cases under the assumption that the all-zero codeword $\underline{c}^{(0)}$ is transmitted.



(a) The error event that the all-zero codeword is not in the list.
(b) The error event that the all-zero codeword is in the list but not the closest one.

Fig. 2. Graphical illustrations of the decoding error events.

Case 1. The all-zero codeword is not in the list $\mathcal{L}_{\underline{y}}$ (see Fig. 2 (a)), that is, $\underline{y} \notin \mathcal{R}$, which means that at least $d^* + 1$ errors occur over the BSC. This probability is

$$\Pr\{\underline{y} \notin \mathcal{R}\} = \sum_{m=d^*+1}^{n} \binom{n}{m} p_b^m (1-p_b)^{n-m}. \tag{12}$$

Case 2. The all-zero codeword is in the list $\mathcal{L}_{\underline{y}}$, but is not the closest one to $\underline{y}$ (see Fig. 2 (b)), which is equivalent to the event $\left\{E, \underline{y} \in \mathcal{R}\right\}$. This probability is upper-bounded by

$$\Pr\left\{E, \underline{y} \in \mathcal{R}\right\} \leq \Pr\left\{ \bigcup_{d \leq 2d^*} E_d, \;\; \underline{y} \in \mathcal{R} \right\} \tag{13}$$

since all codewords in the list $\mathcal{L}_{\underline{y}}$ are at most $2d^*$ away from the all-zero codeword and not all codewords of a specific weight are in the list. The above upper bound involves only truncated weight spectrum. However, the region $\mathcal{R}$ is in unknown shape and may not be symmetric, which causes difficulties when computing the upper bound. To circumvent this difficulty, we

may enlarge $\mathcal{R}$ to $\mathbb{R}^n$ and get

$$\begin{aligned}\Pr\left\{E, \underline{y} \in \mathcal{R}\right\} &\leq \Pr\left\{\bigcup_{d\leq 2d^*} E_d, \ \ \underline{y} \in \mathcal{R}\right\} &(14)\\ &\leq \Pr\left\{\bigcup_{d\leq 2d^*} E_d, \ \ \underline{y} \in \mathbb{R}^n\right\} &(15)\\ &= \Pr\left\{\bigcup_{d\leq 2d^*} E_d\right\} \leq T_u(\mathcal{C}_{2d^*}), &(16)\end{aligned}$$

where $T_u(\mathcal{C}_{2d^*})$ is a computable upper bound on $\Pr\left\{\bigcup_{d\leq 2d^*} E_d\right\}$, which depends only on the sub-code $\mathcal{C}_{2d^*}$ consisting of all codewords with Hamming weight no greater than $2d^*$. It is worth pointing out that, although the sub-code $\mathcal{C}_{2d^*}$ may not be linear, most bounding techniques in [2] can be applied to $\mathcal{C}_{2d^*}$ to get such an upper bound under the assumption that the all-zero codeword is transmitted. Hereafter, we use the notation $\mathcal{C}_t \triangleq \{\underline{c} \in \mathcal{C} \mid W_H(\underline{c}) \leq t\}$.

For convenience, we define

$$B(p, N_t, N_\ell, N_u) \triangleq \sum_{m=N_\ell}^{N_u} \binom{N_t}{m} p^m (1-p)^{N_t-m}. \tag{17}$$

The function $B(p, N_t, N_\ell, N_u)$, which will be used over and over again in this paper, is just the probability that the number of bit-errors occurring in a binary vector of total length N_t, when passing through a BSC with cross error probability p, ranges from N_ℓ to N_u. Note that $B(p, N_t, N_\ell, N_u)$ can be calculated recursively independently of codes.

Combining (12), (16) and (17) with (2), we get an upper bound

$$\Pr\{E\} \leq T_u(\mathcal{C}_{2d^*}) + B(p_b, n, d^*+1, n), \tag{18}$$

where the second term in the RHS is computable without requiring the code structure and the first term depends only on the sub-code $\mathcal{C}_{2d^*}$.

On one hand, similar to the SB [10] and the TSB [11], the proposed upper bound (18) involves only truncated weight spectrum, which is hence helpful when the whole weight spectrum is not computable. On the other hand, if the complete weight spectrum is available, the proposed bounding technique can potentially improve any existing upper bounds.

Proposition 1: Let T_u be an upper-bounding technique. We have

$$\Pr\{E\} \leq \min_{0\leq d^*\leq n} \left\{T_u(\mathcal{C}_{2d^*}) + B(p_b, n, d^*+1, n)\right\}, \tag{19}$$

which delivers an upper bound strictly less than 1 and not looser than any existing upper bounds $T_u(\mathcal{C})$.

Proof: Noting that $T_u(\mathcal{C}_0) = 0$ and $B(p_b, n, 1, n) = 1-(1-p_b)^n$, we have, by setting $d^* = 0$,

$$\Pr\{E\} < 1. \tag{20}$$

Similarly, noting that $T_u(\mathcal{C}_{2n}) = T_u(\mathcal{C})$ and $B(p_b, n, n+1, n) = 0$, we have, by setting $d^* = n$,

$$\Pr\{E\} \leq T_u(\mathcal{C}). \tag{21}$$

■

Taking the conventional union bound as T_u, we have

Theorem 1: Let $d_{\min}$ be the minimum Hamming weight of the code $\mathcal{C}$. We have

$$\Pr\{E\} \leq \min_{0\leq d^*\leq n} \left\{ \sum_{d_{\min}\leq d\leq 2d^*} A_d Q\left(\frac{\sqrt{d}}{\sigma}\right) + B(p_b, n, d^*+1, n) \right\}. \tag{22}$$

Proof: It can be proved by substituting the conventional union bound for $T_u(\mathcal{C}_{2d^*})$ (in the same form as shown in (8)) into (19). ■

Remark. The bound (22), which is slightly different from that proposed in [19], requires higher computational loads than the conventional union bound. The overhead is caused by recursively computing $B(p_b, n, d^*+1, n)$ and minimizing over d^*. If we do not perform the optimization and simply set $d^* = n$, we get the conventional union bound, implying that the technique can potentially improve the conventional union bound, as stated in Proposition 1.

IV. Improved Union Bounds

We have interpreted the "truncated" union bound as an upper-bounding technique based on the GFBT, where the region $\mathcal{R}$ is defined by a sub-optimal decoding algorithm. To bound

$\Pr\{E, \underline{y} \in \mathcal{R}\}$, we have enlarged $\mathcal{R}$ to $\mathbb{R}^n$, as shown in the derivation from (14) to (15). The objective of this section is to reduce the effect of such an enlargement.

Noticing that the event $\underline{y} \in \mathcal{R}$ is equivalent to the event $W_H(\hat{\underline{y}}) \leq d^*$, we have

Proposition 2:

$$\Pr\{E\} \leq \min_{0 \leq d^* \leq n} \left\{ \sum_{d \leq 2d^*} \Pr\left\{E_d, W_H(\hat{\underline{y}}) \leq d^*\right\} + B(p_b, n, d^*+1, n) \right\}. \tag{23}$$

Proof: For any d^* ($0 \leq d^* \leq n$),

$$\begin{aligned} \Pr\{E\} &\leq \Pr\{E, \underline{y} \in \mathcal{R}\} + \Pr\{\underline{y} \notin \mathcal{R}\} \\ &\leq \Pr\left\{ \bigcup_{d \leq 2d^*} E_d, W_H(\hat{\underline{y}}) \leq d^* \right\} + B(p_b, n, d^*+1, n) \\ &\leq \sum_{d \leq 2d^*} \Pr\left\{E_d, W_H(\hat{\underline{y}}) \leq d^*\right\} + B(p_b, n, d^*+1, n). \end{aligned} \tag{24}$$

■

In this section, we focus on how to upper-bound $\Pr\left\{E_d, W_H(\hat{\underline{y}}) \leq d^*\right\}$ for any given d and d^*. Without loss of generality, we assume that $A_d \geq 1$ and denote all the codewords with weight d by $\underline{c}^{(\ell)}$, $1 \leq \ell \leq A_d$. Let $E_{0 \to \ell}$ be the event that $\underline{c}^{(\ell)}$ is nearer than $\underline{c}^{(0)}$ to $\underline{y}$.

A. Union Bounds Using Pair-Wise Error Probability

Lemma 2:

$$\Pr\left\{E_{0\to 1}, W_H(\hat{\underline{y}}) \leq d^*\right\} \leq Q(\sqrt{d}/\sigma) B\left(p_b, n-d, 0, d^*-1\right). \tag{25}$$

Proof: Without loss of generality, let $\underline{c}^{(1)} \triangleq (\underbrace{1 \cdots 1}_{d} \underbrace{0 \cdots 0}_{n-d})$. Denote $\underline{y}_0^{d-1} \triangleq (y_0, \cdots, y_{d-1})$ and $\underline{y}_d^{n-1} \triangleq (y_d, \cdots, y_{n-1})$. Evidently, only $\underline{y}_0^{d-1}$ can cause the decoding error event that $\underline{c}^{(1)}$ is nearer than $\underline{c}^{(0)}$ to $\underline{y}$. In other words, the event $E_{0\to 1}$ is independent of $\underline{y}_d^{n-1}$ and $\Pr\{E_{0\to 1}\} = Q\left(\sqrt{d}/\sigma\right)$. Also notice that the received signal vector $\underline{y}$ which can cause the event $E_{0\to 1}$ must satisfy $W_H(\hat{\underline{y}}_0^{d-1}) \geq 1$. Hence $\left\{\underline{y} | E_{0\to 1}, W_H(\hat{\underline{y}}) \leq d^*\right\} \subseteq \left\{\underline{y} | E_{0\to 1}, W_H(\hat{\underline{y}}_d^{n-1}) \leq d^* - 1\right\}$. Then

we have

$$\begin{aligned}\Pr\left\{E_{0\to 1}, W_H(\underline{\hat{y}}) \le d^*\right\} &\le \Pr\left\{E_{0\to 1}, W_H(\underline{\hat{y}}_d^{n-1}) \le d^*-1\right\} &(26)\\ &= \Pr\left\{E_{0\to 1}\right\}\Pr\left\{W_H(\underline{\hat{y}}_d^{n-1}) \le d^*-1\right\} &(27)\\ &= Q(\sqrt{d}/\sigma)B\left(p_b, n-d, 0, d^*-1\right). &(28)\end{aligned}$$

■

Theorem 2:

$$\Pr\left\{E_d, W_H(\underline{\hat{y}}) \le d^*\right\} \le A_d Q(\sqrt{d}/\sigma)B\left(p_b, n-d, 0, d^*-1\right). \quad (29)$$

Proof: By union bounds and the symmetries of the error events,

$$\begin{aligned}\Pr\left\{E_d, W_H(\underline{\hat{y}}) \le d^*\right\} &= \Pr\left\{\bigcup_{1\le\ell\le A_d} E_{0\to\ell}, W_H(\underline{\hat{y}}) \le d^*\right\} &(30)\\ &\le \sum_{1\le\ell\le A_d}\Pr\left\{E_{0\to\ell}, W_H(\underline{\hat{y}}) \le d^*\right\} &(31)\\ &= A_d\Pr\left\{E_{0\to 1}, W_H(\underline{\hat{y}}) \le d^*\right\} &(32)\\ &\le A_d Q(\sqrt{d}/\sigma)B\left(p_b, n-d, 0, d^*-1\right). &(33)\end{aligned}$$

■

B. Union Bounds Using Triplet-Wise Error Probability

Temporarily, we assume that $A_d \ge 2$ is even. Then we have

$$\Pr\{E_d, W_H(\underline{\hat{y}}) \le d^*\} \le \sum_{1\le\ell\le A_d/2}\Pr\left\{E_{0\to(2\ell-1)}\bigcup E_{0\to 2\ell}, W_H(\underline{\hat{y}}) \le d^*\right\}. \quad (34)$$

If we can find ways to calculate or upper-bound $\Pr\left\{E_{0\to(2\ell-1)}\bigcup E_{0\to 2\ell}, W_H(\underline{\hat{y}}) \le d^*\right\}$, we may improve the conventional union bound.

In this paper, we refer to the probability $\Pr\left\{E_{0\to 1}\bigcup E_{0\to 2}\right\}$ as *triplet-wise error probability*. We have the following lemma.

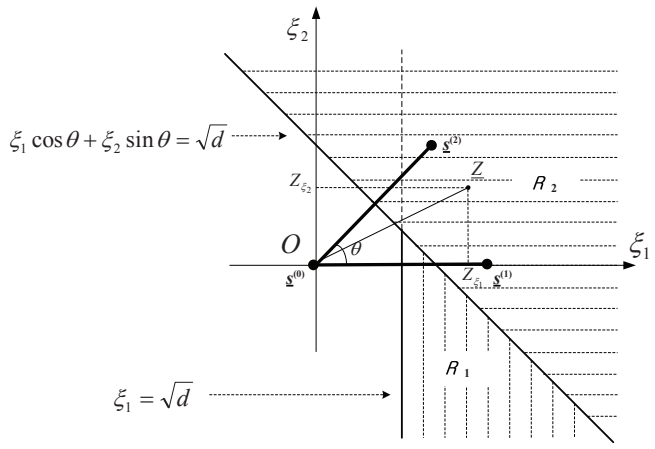


Fig. 3. Geometrical interpretation of the triplet-wise error probability.

Lemma 3: Let $\underline{c}^{(0)}$ be the all-zero codeword with bipolar image $\underline{s}^{(0)}$. Let $\underline{c}^{(1)}$ and $\underline{c}^{(2)}$ be the codewords of Hamming weight d with bipolar images $\underline{s}^{(1)}$ and $\underline{s}^{(2)}$, respectively. The triplet-wise error probability

$$\Pr\left\{E_{0\to1}\bigcup E_{0\to2}\right\}=Q(\sqrt{d}/\sigma)+\int_{\sqrt{d}}^{+\infty}f(\xi_1)\int_{-\infty}^{\frac{\sqrt{d}-\xi_1\cos\theta}{\sin\theta}}f(\xi_2)\ \mathrm{d}\xi_2\ \mathrm{d}\xi_1, \tag{35}$$

where $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-x^2/(2\sigma^2)}$ is the probability density function of $\mathcal{N}(0,\sigma^2)$ and θ is the angle formed by the two vectors $\overrightarrow{\underline{s}^{(0)}\underline{s}^{(1)}}$ and $\overrightarrow{\underline{s}^{(0)}\underline{s}^{(2)}}$. Furthermore, the triplet-wise error probability is a non-decreasing function of θ.

Proof: Similar to the proof of Lemma 1, we have sketched the two vectors $\overrightarrow{\underline{s}^{(0)}\underline{s}^{(1)}}$ and $\overrightarrow{\underline{s}^{(0)}\underline{s}^{(2)}}$ in a two-dimensional space, as shown in Fig. 3, where we have chosen $\underline{s}^{(0)}$ as the origin O and arranged $\overrightarrow{\underline{s}^{(0)}\underline{s}^{(1)}}$ on the abscissa axis $\overrightarrow{O\xi_1}$.

Assume that $\underline{s}^{(0)}$ is transmitted and $\underline{y}=\underline{s}^{(0)}+\underline{z}$ is received, where $\underline{z}$ is a sample from a random vector $\underline{Z}$ whose components are independent and identically distributed as $\mathcal{N}(0,\sigma^2)$. Let Z_{ξ_1} and Z_{ξ_2} be the two independent Gaussian random variables by projecting $\underline{Z}$ onto the abscissa axis and ordinate axis, respectively. Specifically, say, Z_{ξ_1} is the inner product $\langle\underline{Z},\frac{\underline{s}^{(1)}-\underline{s}^{(0)}}{\|\underline{s}^{(1)}-\underline{s}^{(0)}\|}\rangle$. It is well-known that only (Z_{ξ_1},Z_{ξ_2}) can cause the error event $\{E_{0\to1}\bigcup E_{0\to2}\}$. Actually, as shown

in Fig. 3, the error event $\{E_{0\to 1}\bigcup E_{0\to 2}\}$ occurs if and only if the vector (Z_{ξ_1}, Z_{ξ_2}) falls into the shaded region, which can be partitioned into

$$\mathcal{R}_1 \triangleq \left\{(\xi_1,\xi_2)|\xi_1 \geq \sqrt{d},\ \xi_1\cos\theta + \xi_2\sin\theta < \sqrt{d}\right\}, \tag{36}$$

$$\mathcal{R}_2 \triangleq \left\{(\xi_1,\xi_2)|\xi_1\cos\theta + \xi_2\sin\theta \geq \sqrt{d}\right\}. \tag{37}$$

Since $\Pr\{\mathcal{R}_1\} = \int_{\sqrt{d}}^{+\infty} f(\xi_1)\int_{-\infty}^{\frac{\sqrt{d}-\xi_1\cos\theta}{\sin\theta}} f(\xi_2)\ \mathrm{d}\xi_2\ \mathrm{d}\xi_1$ and $\Pr\{\mathcal{R}_2\} = Q(\sqrt{d}/\sigma)$, we have

$$\begin{aligned}\Pr\left\{E_{0\to 1}\bigcup E_{0\to 2}\right\} &= \Pr\{\mathcal{R}_1\} + \Pr\{\mathcal{R}_2\} \\ &= \int_{\sqrt{d}}^{+\infty} f(\xi_1)\int_{-\infty}^{\frac{\sqrt{d}-\xi_1\cos\theta}{\sin\theta}} f(\xi_2)\ \mathrm{d}\xi_2\ \mathrm{d}\xi_1 + Q(\sqrt{d}/\sigma). \end{aligned}\tag{38}$$

To prove the monotonicity, it suffices to prove that $\frac{\sqrt{d}-\xi_1\cos\theta}{\sin\theta}$ increases with θ for $\xi_1 \geq \sqrt{d}$. This can be verified by noting that its derivative $\frac{\xi_1-\sqrt{d}\cos\theta}{\sin^2\theta} \geq 0$ for $\xi_1 \geq \sqrt{d}$. ■

Lemma 4: For any two codewords $\underline{c}^{(1)}$ and $\underline{c}^{(2)}$ of Hamming weight d, the triplet-wise error probability[1]

$$\Pr\left\{E_{0\to 1}\bigcup E_{0\to 2}\right\} \leq 2Q(\sqrt{d}/\sigma) - Q^2(\sqrt{d}/\sigma). \tag{39}$$

Proof: From Lemmas 1 and 3, we can substitute $\theta = \pi/2$ into (35) to complete the proof. ■

Remark. From Lemmas 1 and 3, in the case of $\arccos\sqrt{\frac{d}{n}} < \pi/4$, we may substitute $\theta = 2\arccos\sqrt{\frac{d}{n}}$ into (35) to get a tighter bound, however, which needs higher computational loads.

Lemma 5: For any two codewords $\underline{c}^{(1)}$ and $\underline{c}^{(2)}$ of Hamming weight d,

$$\Pr\left\{E_{0\to 1}\bigcup E_{0\to 2}, W_H(\underline{\hat{y}}) \leq d^*\right\} \leq \left(2Q(\sqrt{d}/\sigma) - Q^2(\sqrt{d}/\sigma)\right) B(p_b, n-2d, 0, d^*-1). \tag{40}$$

Proof: Without loss of generality, assume that

$$\underline{c}^{(1)} \triangleq (c_0^{(1)}\cdots c_{2d-1}^{(1)}\underbrace{0\cdots 0}_{n-2d}) \tag{41}$$

[1]As pointed out by an anonymous reviewer that the RHS of (39) is the same as the symbol error probability of quadrature phase shift keying (QPSK) over AWGN channels [23].

and

$$\underline{c}^{(2)} \stackrel{\Delta}{=} (c_0^{(2)} \cdots c_{2d-1}^{(2)} \underbrace{0 \cdots 0}_{n-2d}). \tag{42}$$

Then only $\underline{y}_0^{2d-1}$ can cause the event that $\underline{c}^{(1)}$ or $\underline{c}^{(2)}$ are nearer than $\underline{c}^{(0)}$ to $\underline{y}$. Also notice that the received signal vector $\underline{y}$ which can cause the event $E_{0\to1} \bigcup E_{0\to2}$ must satisfy $W_H(\hat{\underline{y}}_0^{2d-1}) \geq 1$. Hence $\left\{\underline{y} | E_{0\to1} \bigcup E_{0\to2}, W_H(\hat{\underline{y}}) \leq d^*\right\} \subseteq \left\{\underline{y} | E_{0\to1} \bigcup E_{0\to2}, W_H(\hat{\underline{y}}_{2d}^{n-1}) \leq d^* - 1\right\}$. Then we have

$$\begin{aligned}
\Pr\left\{E_{0\to1} \bigcup E_{0\to2}, W_H(\hat{\underline{y}}) \leq d^*\right\} &\leq \Pr\left\{E_{0\to1} \bigcup E_{0\to2}, W_H(\hat{\underline{y}}_{2d}^{n-1}) \leq d^* - 1\right\} && (43)\\
&= \Pr\left\{E_{0\to1} \bigcup E_{0\to2}\right\} \Pr\left\{W_H(\hat{\underline{y}}_{2d}^{n-1}) \leq d^* - 1\right\} && (44)\\
&\leq \left(2Q(\sqrt{d}/\sigma) - Q^2(\sqrt{d}/\sigma)\right) B(p_b, n-2d, 0, d^*-1) && (45)
\end{aligned}$$

from Lemma 4. ■

The main result of this subsection is the following theorem, which shows that the union bound based on triplet-wise error probabilities can be tighter than the conventional union bound based on pair-wise error probabilities.

Theorem 3: If A_d is even,

$$\Pr\{E_d, W_H(\hat{\underline{y}}) \leq d^*\} \leq A_d \left(Q(\sqrt{d}/\sigma) - \frac{1}{2}Q^2(\sqrt{d}/\sigma)\right) B(p_b, n-2d, 0, d^*-1); \tag{46}$$

if A_d is odd,

$$\begin{aligned}
\Pr\{E_d, W_H(\hat{\underline{y}}) \leq d^*\} \;\leq\; & (A_d - 1)\left(Q(\sqrt{d}/\sigma) - \frac{1}{2}Q^2(\sqrt{d}/\sigma)\right) B(p_b, n-2d, 0, d^*-1)\\
& + Q(\sqrt{d}/\sigma) B(p_b, n-d, 0, d^*-1). \qquad (47)
\end{aligned}$$

Proof: If A_d is even, we have

$$\begin{aligned}
\Pr\{E_d, W_H(\hat{\underline{y}}) \leq d^*\} &\leq \sum_{1\leq\ell\leq A_d/2} \Pr\left\{E_{0\to(2\ell-1)} \bigcup E_{0\to2\ell}, W_H(\hat{\underline{y}}) \leq d^*\right\}\\
&\leq \frac{A_d}{2}\left(2Q(\sqrt{d}/\sigma) - Q^2(\sqrt{d}/\sigma)\right) B(p_b, n-2d, 0, d^*-1)\\
&= A_d\left(Q(\sqrt{d}/\sigma) - \frac{1}{2}Q^2(\sqrt{d}/\sigma)\right) B(p_b, n-2d, 0, d^*-1), \qquad (48)
\end{aligned}$$

which follows from the symmetries of the error events and Lemma 5.

If A_d is odd, we have

$$\begin{aligned}
&\Pr\{E_d, W_H(\underline{\hat{y}}) \le d^*\} \\
&\le \sum_{1\le \ell \le (A_d-1)/2} \Pr\left\{E_{0\to(2\ell-1)} \bigcup E_{0\to 2\ell}, W_H(\underline{\hat{y}}) \le d^*\right\} + \Pr\left\{E_{0\to A_d}, W_H(\underline{\hat{y}}) \le d^*\right\} \quad (49)\\
&\le (A_d-1)\left(Q(\sqrt{d}/\sigma) - \frac{1}{2}Q^2(\sqrt{d}/\sigma)\right) B(p_b, n-2d, 0, d^*-1) \\
&\quad + Q(\sqrt{d}/\sigma) B(p_b, n-d, 0, d^*-1), \quad (50)
\end{aligned}$$

which follows from the symmetries of the error events and Lemmas 2 and 5. ■

Note that the bounds in Theorem 3 will not always improve the bounds in Theorem 2, since it may happen that $B(p_b, n-2d, 0, d^*-1) > B(p_b, n-d, 0, d^*-1)$.

V. Adaptations of the Improved Union Bounds

A. Bounds for An Ensemble of Codes

As we know, most existing bounds are applied to ensembles of codes as well as specific codes. However, the bounds given in Theorem 3 can not be applied directly to ensembles of codes because the average weight spectra of a code ensemble are usually not be integer-valued.

Theorem 4: Consider a code ensemble $\mathscr{C}$ with probability distribution $\Pr\{\mathcal{C}\}$, $\mathcal{C} \in \mathscr{C}$. Let $\{A_d^{\mathcal{C}}\}$ be the weight spectrum of a specific code $\mathcal{C}$. Then $A_d = \sum_{\mathcal{C}} \Pr\{\mathcal{C}\} A_d^{\mathcal{C}}$ is referred to as the average weight spectra. Define

$$h(A_d) \triangleq \min\left\{ \begin{array}{c} A_d Q(\sqrt{d}/\sigma) B(p_b, n-d, 0, d^*-1), \\ (A_d-1)\left(Q(\sqrt{d}/\sigma) - \frac{1}{2}Q^2(\sqrt{d}/\sigma)\right) B(p_b, n-2d, 0, d^*-1) + Q(\sqrt{d}/\sigma) \end{array} \right\}. \quad (51)$$

Then $\Pr\{E_d, W_H(\underline{\hat{y}}) \le d^*\} \le h(A_d)$.

Proof: From Theorem 2, we have

$$\begin{aligned}\Pr\{E_d, W_H(\underline{\hat{y}}) \le d^*\} &= \sum_{\mathcal{C}} \Pr\{\mathcal{C}\}\Pr\{E_d, W_H(\underline{\hat{y}}) \le d^*|\mathcal{C}\} \\ &\le \sum_{\mathcal{C}} \Pr\{\mathcal{C}\} A_d^{\mathcal{C}} Q(\sqrt{d}/\sigma) B(p_b, n-d, 0, d^*-1) \\ &= A_d Q(\sqrt{d}/\sigma) B(p_b, n-d, 0, d^*-1). \end{aligned} \tag{52}$$

It can be verified from Theorem 3 that, for any $A_d^{\mathcal{C}} \ge 0$,

$$\begin{aligned}\Pr\{E_d, W_H(\underline{\hat{y}}) \le d^*|\mathcal{C}\} &\le (A_d^{\mathcal{C}} - 1)\left(Q(\sqrt{d}/\sigma) - \frac{1}{2}Q^2(\sqrt{d}/\sigma)\right) B(p_b, n-2d, 0, d^*-1) \\ &\quad + Q(\sqrt{d}/\sigma). \end{aligned} \tag{53}$$

Then, we have

$$\begin{aligned}&\Pr\{E_d, W_H(\underline{\hat{y}}) \le d^*\} \\ &= \sum_{\mathcal{C}} \Pr\{\mathcal{C}\}\Pr\{E_d, W_H(\underline{\hat{y}}) \le d^*|\mathcal{C}\} \\ &\le \sum_{\mathcal{C}} \Pr\{\mathcal{C}\}\left\{(A_d^{\mathcal{C}} - 1)\left(Q(\sqrt{d}/\sigma) - \frac{1}{2}Q^2(\sqrt{d}/\sigma)\right)B(p_b, n-2d, 0, d^*-1) + Q(\sqrt{d}/\sigma)\right\} \\ &= (A_d - 1)\left(Q(\sqrt{d}/\sigma) - \frac{1}{2}Q^2(\sqrt{d}/\sigma)\right) B(p_b, n-2d, 0, d^*-1) + Q(\sqrt{d}/\sigma). \end{aligned} \tag{54}$$

Combining (52) and (54), and taking into account the definition of $h(A_d)$, we have

$$\Pr\{E_d, W_H(\underline{\hat{y}}) \le d^*\} \le h(A_d). \tag{55}$$

■

We now summarize the main result in the following theorem, which can be applied to both specific codes and ensembles of codes.

Theorem 5: Let $\{A_d\}$ be the (average) weight spectrum of a specific code or a code ensemble. The word-error probability can be upper-bounded by

$$\Pr\{E\} \le \min_{0 \le d^* \le n} \left\{ \sum_{d \le 2d^*} h(A_d) + B(p_b, n, d^*+1, n) \right\}. \tag{56}$$

Proof: Since a specific code is a special case of a code ensemble with a degraded probability distribution, we consider only a code ensemble.

Combining Theorem 4 with Proposition 2, or equivalently, substituting (55) into (23), we then have (56), completing the proof. ■

B. Bounds for Bit-Error Probabilities

In order to adapt the upper bound (56) to the bit-error probability, we define

$$\hat{i}_d \triangleq \max\{i \mid A_{i,d} > 0\}, \tag{57}$$

$$A'_d \triangleq \sum_i \frac{i}{k} A_{i,d} \tag{58}$$

and

$$h'(A_d) \triangleq \min\left\{ \begin{array}{c} A'_d Q(\sqrt{d}/\sigma) B(p_b, n-d, 0, d^*-1), \\ \frac{\hat{i}_d}{k}\left((A_d-1)\left(Q(\sqrt{d}/\sigma) - \frac{1}{2}Q^2(\sqrt{d}/\sigma)\right) B(p_b, n-2d, 0, d^*-1) + Q(\sqrt{d}/\sigma)\right) \end{array} \right\}. \tag{59}$$

We have the following theorem.

Theorem 6: The bit-error probability can be upper-bounded by

$$P_b \leq \min_{0 \leq d^* \leq n} \left\{ \sum_{d \leq 2d^*} h'(A_d) + B(p_b, n, d^*+1, n) \right\}. \tag{60}$$

Proof: Let $\underline{\hat{U}} \in \mathbb{F}_2^k$ be the binary output vector from a decoder when the input to the encoder is $\underline{U}$. The bit-error probability associated with the decoder is defined as [24, p. 9]

$$P_b \triangleq \frac{1}{k} \sum_{0 \leq i \leq k-1} \Pr\{\hat{u}_i \neq u_i\}. \tag{61}$$

Given that the all-zero codeword is transmitted, the bit-error probability can be rewritten as

$$P_b = \mathbf{E}\left\{ \frac{W_H(\underline{\hat{U}})}{k} \right\}, \tag{62}$$

where $\mathbf{E}$ is the mathematical expectation.

Now we assume that Algorithm 1 is implemented as the decoder. Without loss of generality, we make an assumption that $\hat{\underline{U}}$ is uniformly at random chosen from $\mathbb{F}_2^k$ whenever Algorithm 1 reports a decoding error. Recall that $\mathcal{R} = \left\{\underline{y} | \underline{c}^{(0)} \in \mathcal{L}_{\underline{y}}\right\}$ as defined in (11). We assume the following partition $\mathcal{R} = \bigcup_d \mathcal{R}_d$, where $\underline{y} \in \mathcal{R}_d$ if and only if Algorithm 1 outputs one codeword with Hamming weight d. We have

$$\begin{aligned} kP_b &= \Pr\{\underline{y} \in \mathcal{R}\}\mathbf{E}\{W_H(\hat{\underline{U}})|\underline{y} \in \mathcal{R}\} + \Pr\{\underline{y} \notin \mathcal{R}\}\mathbf{E}\{W_H(\hat{\underline{U}})|\underline{y} \notin \mathcal{R}\} \\ &\leq \Pr\{\underline{y} \in \mathcal{R}\}\mathbf{E}\{W_H(\hat{\underline{U}})|\underline{y} \in \mathcal{R}\} + k\Pr\{\underline{y} \notin \mathcal{R}\} \\ &\leq \sum_{d \leq 2d^*} \Pr\{\underline{y} \in \mathcal{R}_d\}\mathbf{E}\{W_H(\hat{\underline{U}})|\underline{y} \in \mathcal{R}_d\} + kB(p_b, n, d^*+1, n), \end{aligned} \tag{63}$$

where we have used the fact that $\mathbf{E}\{W_H(\hat{\underline{U}})|\underline{y} \notin \mathcal{R}\} \leq k$.

Now we focus on how to upper-bound $\Pr\{\underline{y} \in \mathcal{R}_d\}\mathbf{E}\{W_H(\hat{\underline{U}})|\underline{y} \in \mathcal{R}_d\}$ for any given $d \leq 2d^*$.

On one hand,

$$\mathbf{E}\{W_H(\hat{\underline{U}})|\underline{y} \in \mathcal{R}_d\} \leq \hat{i}_d \tag{64}$$

by the definition of $\hat{i}_d$ and

$$\Pr\{\underline{y} \in \mathcal{R}_d\} \leq (A_d - 1)\left(Q(\sqrt{d}/\sigma) - \frac{1}{2}Q^2(\sqrt{d}/\sigma)\right)B(p_b, n-2d, 0, d^*-1) + Q(\sqrt{d}/\sigma) \tag{65}$$

from the unified upper bound (53) based on triplet-wise error probabilities.

On the other hand, we assume the following partition $\mathcal{R}_d = \bigcup_\ell \mathcal{R}_d^{(\ell)}$, where $\underline{y} \in \mathcal{R}_d^{(\ell)}$ whenever Algorithm 1 outputs $\underline{c}^{(\ell)}$, $1 \leq \ell \leq A_d$. Denote by $\underline{u}^{(\ell)}$ the input binary vector to the encoder corresponding to the codeword $\underline{c}^{(\ell)}$. Since $\Pr\{\underline{y} \in \mathcal{R}_d^{(\ell)}\} \leq \Pr\{E_{0\to\ell}, \underline{y} \in \mathcal{R}\}$, we have

$$\begin{aligned} \Pr\{\underline{y} \in \mathcal{R}_d\}\mathbf{E}\{W_H(\hat{\underline{U}})|\underline{y} \in \mathcal{R}_d\} &= \sum_{1\leq\ell\leq A_d} \Pr\{\underline{y} \in \mathcal{R}_d^{(\ell)}\}W_H(\underline{u}^{(\ell)}) && (66) \\ &\leq \sum_{1\leq\ell\leq A_d} \Pr\{E_{0\to\ell}, \underline{y} \in \mathcal{R}\}W_H(\underline{u}^{(\ell)}) && (67) \\ &\leq kA'_dQ\left(\frac{\sqrt{d}}{\sigma}\right)B(p_b, n-d, 0, d^*-1) && (68) \end{aligned}$$

from the definition of A'_d and Lemma 2.

Now we have two upper bounds on $\Pr\{\underline{y} \in \mathcal{R}_d\}\mathbf{E}\{W_H(\underline{\hat{U}})|\underline{y} \in \mathcal{R}_d\}$. One is (68), and the other can be obtained by combining (64) and (65). Taking into account the definition of $h'(A_d)$, we have

$$\Pr\{\underline{y} \in \mathcal{R}_d\}\mathbf{E}\{W_H(\underline{\hat{U}})|\underline{y} \in \mathcal{R}_d\} \leq kh'(A_d). \tag{69}$$

Substituting (69) into (63) and minimizing over d^*, we have

$$kP_b \leq \min_{0 \leq d^* \leq n} \left\{ \sum_{d \leq 2d^*} kh'(A_d) + kB(p_b, n, d^* + 1, n) \right\}. \tag{70}$$

Dividing by k on the both sides of (70), we complete the proof. ■

Remark. The bound on the bit-error probability given above is applicable to the optimal decoding algorithm that minimizes the bit-error probability, but will not always be applied to the ML decoding algorithm. In other words, the ML decoding algorithm, which is not optimal for minimizing the bit-error probability, may have a higher bit-error probability.

VI. Numerical Results

In this section, by an $[n, k]$ *random linear code*, we mean a code ensemble in which each code is defined by a uniformly at random selected full-rank parity-check matrix of size $(n-k) \times n$. As shown in [25, Appendix D], the average weight spectra of a random linear code $[n, k]$ can be found as

$$A_d = \begin{cases} \binom{n}{d} \frac{2^k - 1}{2^n - 1}, & 0 < d \leq n \\ 1, & d = 0 \end{cases}. \tag{71}$$

We also need to point out that the weight spectra of the compared BCH codes can be found in [26].

A. Comparisons Between the Proposed Bounds and the Existing Bounds

In this subsection, we present four examples to compare the proposed bounds (56) with the existing bounds on word-error probability.

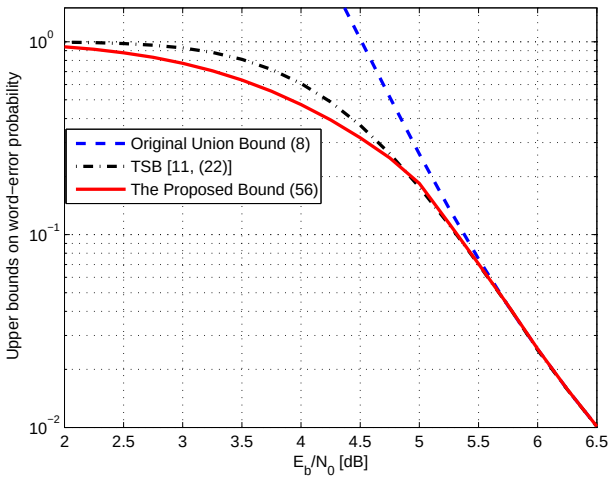


Fig. 4. Comparison between the upper bounds on the word-error probability under ML decoding of random binary linear block codes $[100, 95]$. The compared bounds are the original union bound, the TSB and the proposed bound.

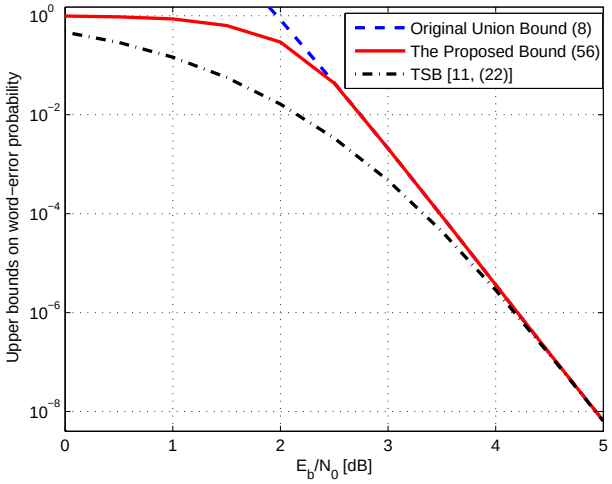


Fig. 5. Comparison between the upper bounds on the word-error probability under ML decoding of random binary linear block codes $[100, 50]$. The compared bounds are the original union bound, the TSB and the proposed bound.

Fig. 4 and Fig. 5 show the comparisons between the original union bound (8), the TSB [11, (22)] and the proposed bound (56) on word-error probability of $[100, 95]$ and $[100, 50]$ random linear codes, respectively, where the former has been used as an example in [2]. The proposed bounds are obtained by optimizing the parameter d^*, which may be varied with SNRs. We can

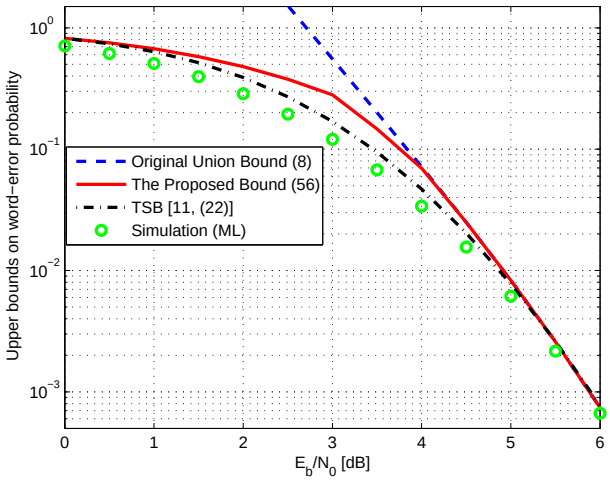


Fig. 6. Comparison between the upper bounds on the word-error probability under ML decoding of BCH code $[31, 26]$. The compared bounds are the original union bound, the TSB and the proposed bound, which are also compared with the ML simulation results.

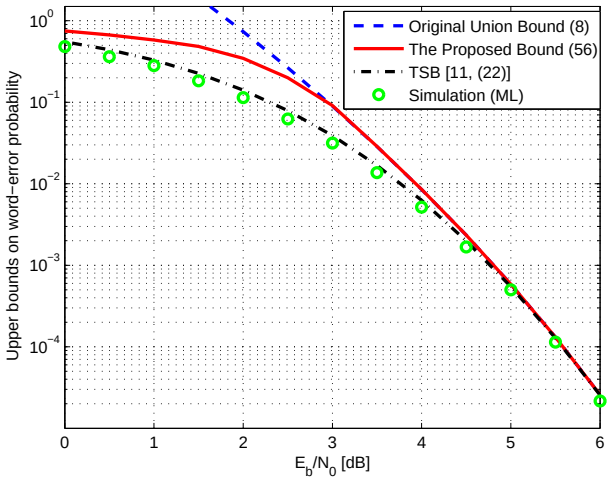


Fig. 7. Comparison between the upper bounds on the word-error probability under ML decoding of BCH code $[31, 21]$. The compared bounds are the original union bound, the TSB and the proposed bound, which are also compared with the ML simulation results.

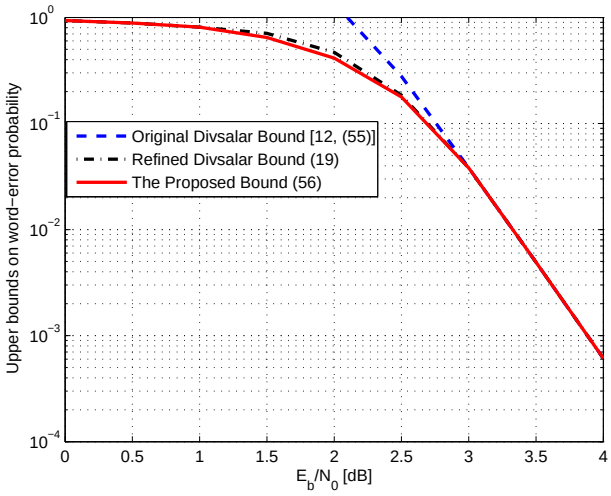


Fig. 8. Comparison between the upper bounds on the word-error probability under ML decoding of BCH code $[63, 39]$. The compared bounds are the original Divsalar bound, the refined Divsalar bound and the proposed bound.

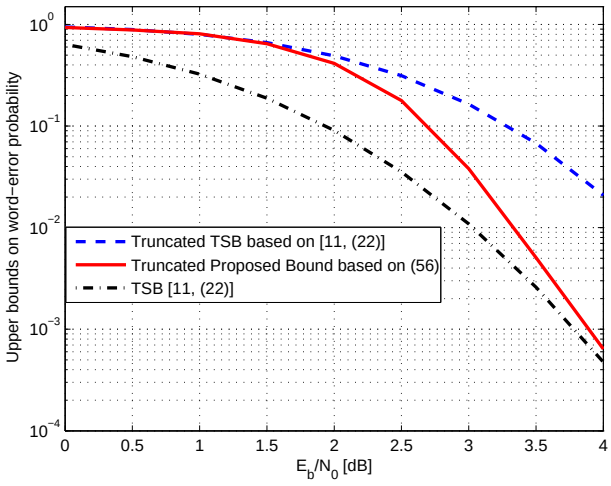


Fig. 9. Comparison between the upper bounds on the word-error probability under ML decoding of BCH code $[63, 39]$. The compared bounds are the truncated TSB, the truncated proposed bound and the TSB. These truncated bounds depend only on the sub-code $\mathcal{C}_{20}$ consisting of all codewords with Hamming weight no greater than 20.

see that the proposed bound improves the original union bound. We can also see that, for the random code $[100, 95]$, the proposed bound is tighter than the TSB in the low-SNR region; while for the random code $[100, 50]$, the proposed bound is looser than the TSB. This coincides with the computational results in [27, Fig. 3], which tells us that the TSB becomes looser in terms of the error exponent with increasing code rates. Note that the solid curve in Fig. 4 is better than that in [28, Fig. 3], since Theorem 4 here improves [28, Theorem 2] by employing the independence between the error events and certain components of the received random vectors.

Fig. 6 and Fig. 7 show the comparisons between the original union bound (8), the TSB [11, (22)] and the proposed bound (56) on word-error probability of $[31, 26]$ and $[31, 21]$ BCH codes, respectively. Also shown are the simulation results. We can see that the proposed bound improves the original union bound especially in the low-SNR region. We can also see that the proposed bound is almost as tight as the TSB for the $[31, 26]$ BCH code but looser than the TSB for the [31, 21] BCH code, which again coincides with the conclusions in [27].

B. Combination of the Proposed Technique with the Existing Bounds

By Proposition 1, we know that the proposed bounding technique can potentially improve any existing upper bounds. To illustrate this, we give an example. Fig. 8 shows the comparisons between the original Divsalar bound [12, (55)], the *refined* Divsalar bound (19) by taking Divsalar bound as T_u and the proposed bound (56) on word-error probability of $[63, 39]$ BCH code, which has been used as an example in [11]. We can see that the refined Divsalar bound improves the original Divsalar bound especially in the low-SNR region. We can also see that the proposed bound (56) is slightly tighter than the refined Divsalar bound. For this $[63, 39]$ BCH code, we have also combined the proposed bounding technique with the SB and the TSB. However, we found that the optimal parameter d^* is n and hence no improvement is achieved for the SB and the TSB.

C. Comparisons Between the Truncated Proposed Bound and the Truncated Existing Bounds

As we have mentioned above Proposition 1, the proposed bounding technique is helpful when the whole weight spectrum is unknown or not computable, as is similar to the SB and the TSB. Hence, it makes sense to compare these truncated bounds. To illustrate this, we take the $[63, 39]$ BCH code as an example. To get the weight spectra, one may need to perform the algorithms in [29]. Given d, the upper bounds of the computational complexity for computing A_d can be found in [29, Lemmas 5 & 7]. For example, one needs about 10^5 and 10^8 attempts of Algorithm 1 in [29] for $d = 9$ and $d = 13$, respectively, as given in [29, Section VI]. Evidently, the fewer A_d ($0 < d \leq n$) we use, the lower computational complexity the algorithm has. Assume that we know only the truncated weight spectrum $\{A_d, d \leq 20\}$. Then we can obtain the truncated proposed bound based on (56) and the truncated TSB based on [11, (22)], as shown in Fig. 9. Also shown in Fig. 9 is the TSB [11, (22)] with the whole weight spectrum. We can see that the truncated proposed bound is looser than the TSB, but tighter than the truncated TSB especially in the high-SNR region. Note that both two truncated bounds are optimized based on the truncated spectrum. For example, the truncated proposed bound is obtained by optimizing the parameter d^* ($0 \leq d^* \leq 10$) in (56).

VII. Conclusions

In this paper, we have presented new techniques to improve the conventional union bounds within the framework of GFBT. Compared with the conventional union bound, the proposed bounds are tighter but have a similar complexity because they involve only the weight spectra and the Q-function. The proposed bounds are also helpful when the whole weight spectrum is unknown or not computable. Numerical results show that the proposed bounds can even improve the TSB in the high-rate region.

ACKNOWLEDGMENT

The authors would like to thank X. Huang and Q.-T. Zhuang for their help. They also would like to thank Prof. Sason for his comments while this work was partially presented in ISIT'2011. They also would like to thank the Associate Editor and the anonymous reviewers for their valuable comments.

REFERENCES

[1] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Transactions on Information Theory*, vol. 48, pp. 3029–3051, December 2002.

[2] I. Sason and S. Shamai, "Performance analysis of linear codes under maximum-likelihood decoding: A tutorial," in *Foundations and Trends in Communications and Information Theory*. Delft, The Netherlands: NOW, July 2006, vol. 3, no. 1-2, pp. 1–225.

[3] T. M. Duman and M. Salehi, "New performance bounds for turbo codes," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 717–723, June 1998.

[4] T. M. Duman, "Turbo codes and turbo coded modulation systems: Analysis and performance bounds," Ph.D. dissertation, Elect. Comput. Eng. Dept., Northeastern Univ., Boston, MA, May 1998.

[5] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 2101–2104, September 1999.

[6] M. Twitto, I. Sason, and S. Shamai, "Tightened upper bounds on the ML decoding error probability of binary linear block codes," *IEEE Transactions on Information Theory*, vol. 53, pp. 1495–1510, April 2007.

[7] E. R. Berlekamp, "The technology of error correction codes," *Proceedings of the IEEE*, vol. 68, pp. 564–593, May 1980.

[8] T. Kasami, T. Fujiwara, T. Takata, K. Tomita, and S. Lin, "Evaluation of the block error probability of block modulation codes by the maximum-likelihood decoding for an AWGN channel," in *Proc. of the 15th Symposium on Information Theory and Its Applications*, Minakami, Japan, September 1992.

[9] T. Kasami, T. Fujiwara, T. Takata, and S. Lin, "Evaluation of the block error probability of block modulation codes by the maximum-likelihood decoding for an AWGN channel," in *Proc. 1993 IEEE Int. Symp. Inform. Theory*, January 1993, p. 68.

[10] H. Herzberg and G. Poltyrev, "Techniques of bounding the probability of decoding error for block coded modulation structures," *IEEE Transactions on Information Theory*, vol. 40, pp. 903–911, May 1994.

[11] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Transactions on Information Theory*, vol. 40, pp. 1284–1292, July 1994.

[12] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," in *Proc. 1999 IEEE Communication Theory Workshop*, Aptos, CA, May 1999.

[13] I. Sason and S. Shamai, "Improved upper bounds on the ML decoding error probability of parallel and serial concatenated turbo codes via their ensemble distance spectrum," *IEEE Transactions on Information Theory*, vol. 46, pp. 24–47, January 2000.

[14] J. Zangl and R. Herzog, "Improved tangential sphere bound on the bit error probability of concatenated codes," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 825–830, May 2001.

[15] D. Divsalar and E. Biglieri, "Upper bounds to error probabilities of coded systems beyond the cutoff rate," *IEEE Trans. Commun.*, vol. 51, no. 12, pp. 2011–2018, December 2003.

[16] S. Yousefi and A. K. Khandani, "Generalized tangential sphere bound on the ML decoding error probability of linear binary block codes in AWGN interference," *IEEE Transactions on Information Theory*, vol. 50, pp. 2810–2815, Novomber 2004.

[17] ——, "A new upper bound on the ML decoding error probability of linear binary block codes in AWGN interference," *IEEE Transactions on Information Theory*, vol. 50, pp. 3026–3036, Novomber 2004.

[18] A. Mehrabian and S. Yousefi, "Improved tangential sphere bound on the ML decoding error probability of linear binary block codes in AWGN and block fading channels," *IEE Proc. Commun.*, vol. 153, pp. 885–893, December 2006.

[19] X. Ma, C. Li, and B. Bai, "Maximum likelihood decoding analysis of LT codes over AWGN channels," in *Proc. of the 6th International Symposium on Turbo Codes and Iterative Information Processing*, Brest, France, September 2010.

[20] E. Agrell, "On the Voronoi neighbor ratio for binary linear block codes," *IEEE Transactions on Information Theory*, vol. 44, pp. 3064–3072, Novomber 1998.

[21] E. Hof, I. Sason, and S. Shamai, "Performance bounds for erasure, list and feedback schemes with linear block codes," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3754–3778, August 2010.

[22] E. Agrell, "Voronoi regions for binary linear block codes," *IEEE Transactions on Information Theory*, vol. 42,

pp. 310–316, January 1996.

[23] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

[24] W. Ryan and S. Lin, *Channel Codes: Classical and Modern*. Cambridge, England: Cambridge University Press, 2009.

[25] M. Twitto, "Tightened upper bounds on the ML decoding error probability of binary linear block codes and applications," Master's thesis, Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel, April 2006.

[26] M. Terada, J. Asatani, and T. Koumoto, "Web site on the weight distribution of BCH and Reed-Muller codes," Available at http://www.infsys.cne.okayama-u.ac.jp/ kusaka/wd/index.html.

[27] M. Twitto and Sason, "On the error exponents of improved tangential sphere bounds," *IEEE Transactions on Information Theory*, vol. 53, pp. 1196–1210, March 2007.

[28] X. Ma, J. Liu, and B. Bai, "New techniques for upper-bounding the MLD performance of binary linear codes," in *Proc. 2011 IEEE Int. Symp. Inform. Theory*, Saint-Petersburg, Russian Federation, August 2011, pp. 2910–2914.

[29] A. M. Barg and I. I. Dumer, "On computing the weight spectrum of cyclic codes," *IEEE Transactions on Information Theory*, vol. 38, pp. 1382–1386, July 1992.